# ANALYSIS OF INCOME INEQUALITY OF EMPLOYEES IN THE SLOVAK REPUBLIC

**Ľubica Sipková – Juraj Sipko**

**Abstract**

The intention of this article is to characterize the inequality of income distribution of employees in the Slovak Republic. The assumption of the applied parametric approach to the calculation of inequality measures is that wages of employees of the Slovak Republic are a result of stochastic processes based on a reasonably smooth probability distribution. Another assumption is that the applied model correctly analytically describes this distribution and therefore also the basic properties and structure of the empirical distribution of wages.

Presented are the results of the analysis of various inequality measures evaluated in the model framework. Various approaches to probability modeling of wages are used while goodness-of-fit has the highest priority when choosing models of the distribution.

The linearity of the pattern of points in a quantile –quantile plot (Q-Q plot) is a visualization of the goodness-of-fit line. The model chosen to interpret the data influenced the calculated single inequality statistic. A feature of the Q-Q plot constructed by displaying the order statistics is that the inequality measures “can be obtained as least squares estimates of a Q-Q diagram”. Tarsitano’s applied approach is compared to the classical approach to the calculation of inequality measures of wages of employees in the Slovak Republic.

**Key words:** income inequality, gross wages, employees, quantiles, probability models

**JEL Code:** J31, C14, C46

## Introduction

The main objectives in income modeling, such as transformation of model components instead of data, opportunity to use the same analytical form of models in regional analyses, functional analytical forms for future relative inequality analyses and application of robust estimate methods with respect to the failure of the standard parametric methods assumptions, could be fulfilled by a quantile approach to distribution modeling thanks to the high flexibility of conceptual quantile models, inequality measures defined by inverses of cdfs (rankits) and robust estimation methods that treat all values equally without requirements of the normality assumption in small area estimation.

Relative income inequality figures based on each region’s income distribution, i. e. depending on its shape, for a fair inequality comparison among regions, the same flexible analytical form of model, should be applied. Besides broad use of central tendency as a measure of the center of differences between all income[1] there is a reasonable phenomenon of grooving utilization of methods based on Galton's quantile modeling techniques and Order statistics theory altogether with semi or nonparametric methods in income studies.

The insistent application of quantile based methods demonstrates their usefulness, especially according to the results which were received by applying them on complex asymmetric distributions with complicated shapes and existence of extremes. The constructed conceptual quantile models treat extremes themselves. Moreover, the future study of extremes could continue with Monte Carlo simulations through the conceptual quantile income models of whole income distribution, and for its subgroups as well, by methods of simulation of extremes by quantile functions.

The four-parameter Kappa generalized quantile distribution is flexible enough for regional - comparative population modeling, which is functional for inequality measures based on its rankits. But prior to inequality evaluation based on the four-parameter Kappa quantile models, it was necessary to establish an appropriate measure of economic inequality.

Tarsitano suggests that his approach "derives simultaneously the law of income distribution and the measure of economic inequality." "In case of a good fit, the least squares estimate of the slope can measure the inequality in the observed incomes."

## 1 Quantile approach to the income probability modeling

Recognizing the same appropriate flexible analytical form of population's distributions makes it possible to describe them fully and provides all the information needed to compare the structures in different geographical regions. It is an advantage when the same type of model can be simply updated and validated for the compared regions, and the estimated parameters of a given analytical form can be used for the whole populations as well as for individual elective samples of the compared populations. Determining the functional form of the population's income distributions and their structures paves the way for comparative parametrical analyses and provides an appropriate approach for assessing differences and similarities in compared income distributions when doing social analyses.

[1] Even though not all values have the same importance in special poverty studies, a moment based description of variability and asymmetry using all values, giving extra power to high extremes, concentration on indicators of economic inequality or poverty with acceptance of its same value for different shapes of income distributions and normal approximations for parametric small area estimation methods.

According to the theory presented by Amartya Sen, (1997) a model of income distribution should be defined in a functional analytical form for further relative inequality analyses, as well as defined by a probability function with good fit for the whole population and for chosen subpopulations. Nowadays, it is often impossible to use the well-known simple income models of the past to express the empirical income distributions of recent more complex populations.

Most of the recently devised econometric models based on a logistic-empirical approach make use of mathematical modifications based probability-generating functions and transformation functions. These intricate econometric models of income are often generalized forms of distribution functions.

The standard statistical approach considers income distribution as the outcome of a stochastic process. The process of modeling probable income distributions using various transformations of input data and resulting in the application of probability distribution function (pdf) driven from known system of probability distribution, have often failed, mainly due to insufficient modeling of income distribution tails. The most frequently used type of a theoretical income model, three-parameter version of logarithm-normal distribution, fits well only in the wide central area of income distribution in the Slovak Republic, but for the marginal regions, other composite types could be more suitable.

The selected atypical approach in the article to define the analytical form principally relies on Gastwirth (1971) standard definition of Lorenz curve, but in its more general form, which uses a quantile income distribution function (qf), i.e. inverse distribution function of income distribution. If $F$ (*cdf*) indicates the proportion of the population $F(x)$ with income no greater than $x$, “and $F^{-1}$ is its inverse (or generalized inverse if $F$ has jumps), the Lorenz curve of $F$ is defined for $0 \leq p \leq 1$ as [Sen 1997, p.139]:

$$L(p) = \int_0^p F^{-1}(p)\,\mathrm{d}p \Big/ \mu \text{, where } \mu = \int_0^1 F^{-1}(p)\,\mathrm{d}p \text{ is the mean of } F\text{”}$$

In addition, the most of commonly presented measures of relative inequality, invariant to changes in population size, mean independent, satisfying symmetry, etc., could be expressed by *qf*. In the stochastic dominance “generalized Lorenz curves are constructed by integrating the inverses of *cdf*s”, i.e. *qf*s, “along the population axis. A simple change of variable converts the integral condition into generalized Lorenz dominance. See Foster and Shorrocks” (Sen 1997).

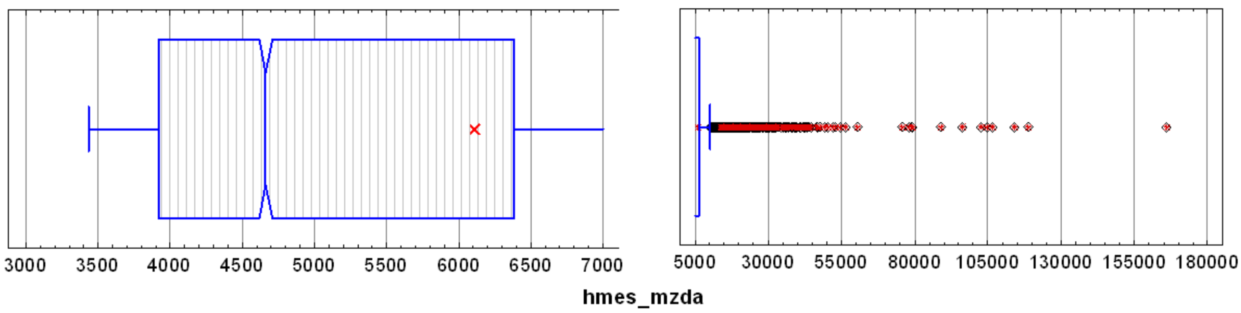
The relatively new field of statistical science called "modeling with quantile functions" has been widely recognized as a significant advance, and is being applied in many areas. Quantile functions make it possible to create a distribution model by combining a number of component models. The quantile-based models provide the flexible approach needed to obtain well-fitted tails in distributional models. This creates new possibilities for probability modeling in cases where the standard approach fails to produce adequate results.

Two potentials are elaborated how to define qf accurately to data properties. The first one, the generalized quantile model fitting may possibly result in definition of the exact shape of high elastic generalized lamda distribution defined by Ramberg and Schmeiser (RSGLD) (Ramberg, Schmeiser 1974) or generalized Kappa distribution (Hosking 1994). Modeling by the four-parameter asymmetrical highly flexible quantile function is a challenging problem because of its two shape parameters in exponents, particularly in the case of more complex income distributions. The results of modeling by generalized Kappa are presented in the third chapter of this article.

More precise results could be achieved by building a conceptual quantile model to reflect data properties, as well as to be relevant to the particular application area. The "construction kit" for the quantile modeling process, in general, analysis of methods and techniques for identifying, estimating and validating probability quantile models especially for skewed heavy left income distributions could be found in Gilchrist (2000).

The procedure and the results of the actual applications of the referred quantile modeling methodology on 1%-random selected sample of non-aggregated official 2010 wages data (hmes_mzda) provided by Organization Trexima, s.r.o for research purposes of employees of the Slovak Republic as well as for its eight NUTS III regions are explained. (The empirical distribution is presented by Box-plot in Fig. 1)

**Fig. 1: Box-plot of average monthly real gross wages of Slovak employees in year 2010 (in euros)**



Source: Author's calculations

It is one of the main goals of this article to apply the quantile probability method of modeling of the whole income distribution for comparison of Slovak regions and for exploration of its potential uses for regional income inequality comparison. The investigation will elaborate and apply the quantile modeling methods that are broadly described in Warren Gilchrist's: Modelling With Quantile Distribution Functions (Gilchrist 2000) and also in Agostiano Tarsitano's: A new Q-Q plot and its application to income data[2].

## 1.1 Solution of Modeling with *pdf*s

Complementary, three-parameter logarithm-normal maximum likelihood estimate was found as log-likelihood and KS best fit income model of population and chosen subpopulations from all 38 well-known two/three-parameter continuous theoretical pdfs offered by statistical software Statgraphics Centurion XVI (Table 1).

**Tab. 1: Results of modeling of wages by well-known *pdf*s**

| Distribution | Est. Parameters | Log Likelihood | Chi-Squared P | KS D | V | U^2 |
|---|---|---|---|---|---|---|
| Loglogistic (3-Parameter) | 3 | -72455.0 | 0.0 | 0.0231775 | 0,0373388 | 0,881959 |
| Loglogistic | 2 | -72467.7 | 0.0 | 0.0252777 | 0,0406725 | 1,14245 |
| Lognormal (3-Parameter) | 3 | -72819.6 | 0.0 | 0.0484607 | 0,0846653 | 6,94377 |
| Lognormal | 2 | -72819.9 | 0.0 | 0.048704 | 0,0849686 | 6,98331 |
| Generalized Logistic | 3 | -73675.8 | 0.0 | 0.0776236 | 0,115476 | 12,5034 |
| Gamma (3-Parameter) | 3 | -73845.8 | 0.0 | 0.0954851 | 0,161453 | 25,1649 |
| Gamma | 2 | -73949.7 | 0.0 | 0.0991269 | 0,167121 | 26,9729 |
| Weibull | 2 | -75251.0 | 0.0 | 0.159989 | 0,27386 | 84,0609 |

Source: Author's calculations

Nevertheless, according to the results of quantitative and graphical verification, the logarithm-normal model does not represent a good approximation of complicated multimodal heavy left and long right tailed total average monthly wages distribution in the Slovak Republic for 2010 and its regions either. The conclusion of classical approach to modeling of hmes_mzda is that none of all applied probability distribution functions (pdfs) fits the whole empirical income data interval well without data transformation and the best fit of applied pdfs represents Logarithm-normal 3-parameter, Gamma 3-parameter and Weibull distribution (see correlation coefficients in Tab. 2), but distributions do not fit the long right tail well enough.

[2] Available on the internet:
http://www3.unisi.it/eventi/GiniLorenz05/ABSTRACT_PAPER_24%20May/PAPER_Tarsitano.pdf

## 1.2 Conceptual quantile modeling based on exploration of suitable *pdf*s

Unlike the standard approach to probability modeling, application of quantile based conceptual modeling methods with its foundation in the Order statistics theory (see e. g. Arnold, Balakrishnan, Nagaraja, 1992) could bring satisfying results because of higher flexibility of quantile functions.

The semi-linear gamma-Pareto type is the composition of gamma and Pareto distributions, without reference to position or scale in its basic inverse forms. The composition by addition was generalized to include position and scale parameters by using the transformation:

$$Q(p) = \alpha + v_1 GAMAINV(p) + v_2 PARETOINV(p),$$

where parameters ν1 and ν2 were set as the functions of probability *p*:

$$v_1 = \omega(1-p) \text{ and } v_2 = \kappa p$$

to give proper weight and spread to gamma and Pareto shapes on each side of the distribution.

The basic form of PARETOINV(*p*) was received from:

$S\ (p) = p$ by the reciprocity rule to:

$R[S(p)] = \dfrac{1}{S(1-p)}$ and by the *Q*-transformation rule to:

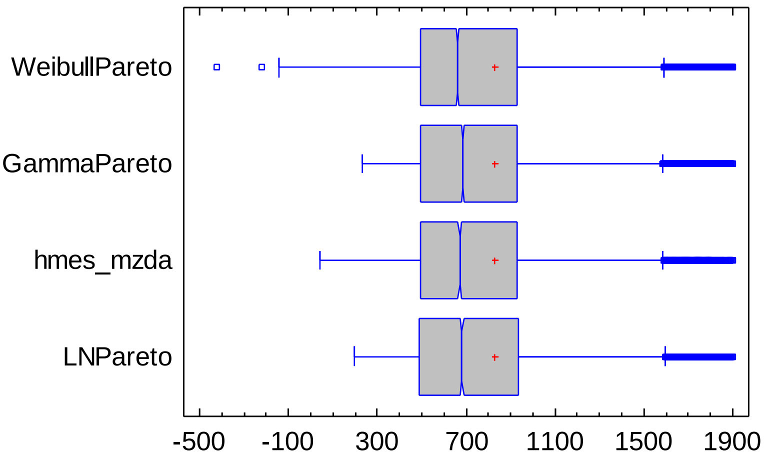
$$S_{PAR}(p) = \frac{1}{(1-p)^{\delta}}, \text{ where } \delta > 0.$$

The composite semi-linear quantile gamma-Pareto model (for LN-Pareto Logarithm-normal inverse was used) for hmes_mzda was built by using modification rules for *qf*s to the final shape:

$$Q(p) = \alpha + \left( \omega(1-p)\, GAMAINV\ (p; \beta.\gamma) + \kappa p \big/ (1-p)^{\delta} \right),\ 0 < p < 1, \delta > 0$$

Both the method of least squares and the method of least absolutes (Gilchrist, 2000) was used to order values of empirical wages, thus $x_{(r)}$ in the regions. The steps of minimization of criterion depend on the general numerical minimization procedure. All the calculations in the process of estimation were done by Solver in Excel. The fitted models were compared graphically (see e.g. Fig. 2) and numerically (see e.g. Tab. 2).

**Fig. 2: Box-plots of distributions**




Source: Author’s calculations

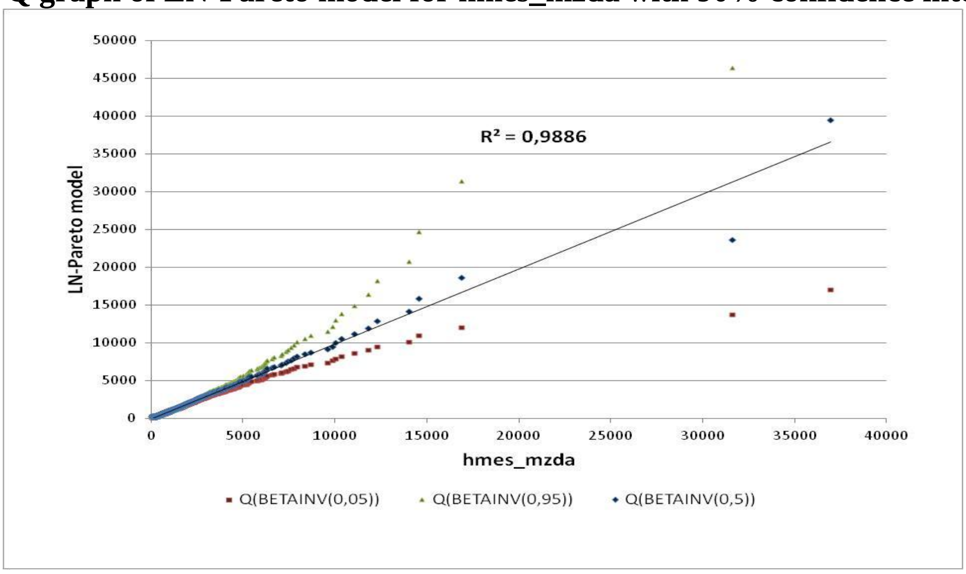
**Tab. 2: Correlation coeficients of wages models**

| | *Loglogistic (3-Parameter)* | *Lognormal (3-Parameter)* | *Gamma (3-Parameter)* | *Weibull* | *GammaPareto* | *WeibullPareto* | *LNPareto* |
|---|---|---|---|---|---|---|---|
| **Correlations** | 0,7829 | 0,8174 | 0,7461 | 0,7438 | 0,9942 | 0,9941 | 0,9943 |
| **$R^2$** | 0,612932 | 0,668143 | 0,556665 | 0,553238 | 0,988434 | 0,988235 | 0,988632 |

Source: Author’s calculations.

Graphical validation of estimated best fitted LNPareto model represented in Fig. 3.

**Fig. 3: Q-Q graph of LN-Pareto model for hmes_mzda with 90% confidence intervals**




Source: Author’s calculations

The applications of LNPareto distribution for the eight NUTS III Slovak regions showed insufficient flexibility. The calculation of appropriate inequality measures on the basis of this respective quantile shape has also proven problematic. We could not reached the same functional form with the same estimation techniques for all the Slovak regions.

## 2. Modeling by generalized Kappa distribution

Fitting the extremely elastic generalized form RS GLD to wages distributions was not successful. Future investigation of techniques for more precise estimation methods, or potential studies of the suitable areas in the four regions of RSGLD two shape parameters in exponents for the typical income shapes, are needed.

The four-parameter Kappa distribution by Hosking (1994), collaborated as a linear-exponential transformation of Gumbel-Exponential shape is defined by its quantile function:

$$Q(p,\lambda) = \lambda_1 + \frac{\lambda_2}{\lambda_3}\left[1 - \left(\frac{1-p^{\lambda_4}}{\lambda_4}\right)^{\lambda_3}\right]; \quad 0 < p \le 1, \lambda_2 > 0 \tag{1}$$

We applied its reduced version in standardized form without a location parameter $\lambda_1$ and a scale parameter $\lambda_2$, with only two exponential parameters $\lambda_3, \lambda_4$, that determine the shape of the basic form of the quantile function. Calculation of the expected value of order statistics follows the Dagum unimodal distribution of the first type, a special case of the four-parameter Kappa distribution, where $\lambda_3 < 0, \lambda_4 < 0$ according to the equation (Tarsitano, p. 12):

$$E(X_{i:n}) = w_{i:n} = \left[p_i^{-\lambda_4} - 1\right]^{-\lambda_3}, i = 1, 2 \ldots n \tag{2}$$

where $p_i$ is the plotting position suggested by Landwehr (with 0.35 and 0.65 proportions) calculated as: $p_i = \frac{i - 0.35}{n}$.

The resulted values of the parameters $\lambda_3, \lambda_4$ were estimated by a downhill simplex maximization of the correlation coefficient (Tarsitano, p. 19):

$$\rho(\lambda_3,\lambda_4) = \frac{\sum_{i=1}^{n}\left\{\left[p_i^{-\lambda_4} - 1\right]^{-\lambda_3} - w_n(\lambda_3,\lambda_4)\right\}X_{i:n}}{\sqrt{S_x b_n(\lambda_3,\lambda_4)}}, i = 1, 2 \ldots n \tag{3}$$

where: $X_{i:n}; i = 1, 2 \ldots n$ is an ordered sample of wages of size $n$,

$$w_n = n^{-1}\sum_{i=1}^{n} w_{i:n},\ i = 1,\ 2\ ...n;\quad S_x = \sum (X_{i:n} - \mu_n)^2 \text{ and } \quad b_n(\lambda_3, \lambda_4) = \sum_{i=1}^{n} (w_{i:n} - w_n)^2 .$$

The results of the two shape parameters $\lambda_3, \lambda_4$ of generalized Kappa distributions calculated for employees of Slovak Republic as well as for its eight NUTS III regions could be found in Fig. 5 and are summarized in Tab. 5.

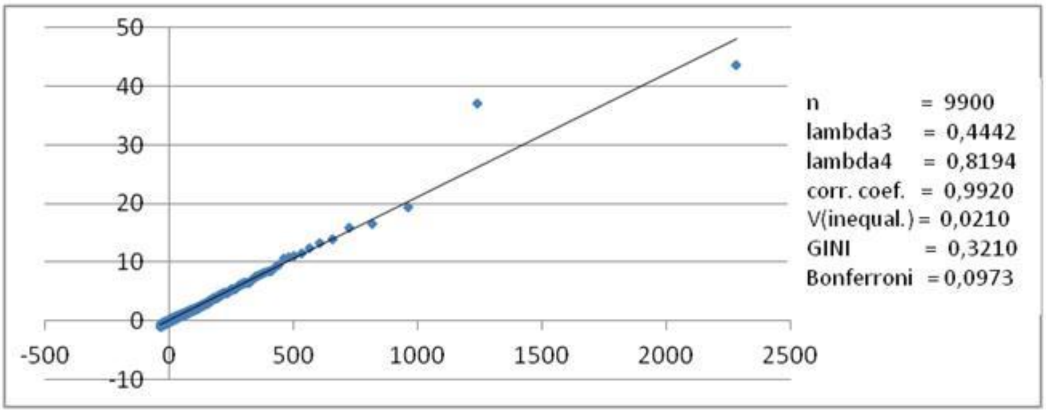
## 2.1 Tarsitano's inequality measures based on rankits in Q-Q plots

To condense information on inequality of wages of employees of the SR and its regions, all 9900 individual observations were retained in calculations of Tarsitano's inequality measures based on rankits in Q-Q plots (see Fig.5) using the equation (Tarsitano, p. 14):

$$V_n(\lambda_3, \lambda_4) = \sum_{i=1}^{n} \left[ \frac{\left[p_i^{-\lambda_4} - 1\right]^{-\lambda_3} - w_n(\lambda_3, \lambda_4)}{\left[p_n^{-\lambda_4} - 1\right]^{-\lambda_3} - w_n(\lambda_3, \lambda_4)} \right] \frac{X_{i:n}}{T},\ i = 1,\ 2\ ...n,\ T = n\mu_n\ ;\quad w_n = \frac{\sum_{i=1}^{n} w_i}{n} \qquad (4)$$

The values of inequality measures are sumarized in the Tab. 5.

**Fig. 5: Q-Q plot for wages of the employees of the Slovak Republic**




**Tab. 5: Results of inequality measures of regions in the Slovak Republic**

| Regions | *n* | lambda3 | lambda4 | Corr. coef. | V(inequal.) | GINI | De Vergottini |
|---|---|---|---|---|---|---|---|
| **Slovak Republic** | 9900 | 0.4442 | 0.8194 | 0.9920 | 0.0210 | 0.3210 | 0.0973 |
| **Region 1** | 2188 | 0.4700 | 0.0655 | 0.9865 | 0.0466 | 0.3808 | 0.1480 |
| **Region 2** | 1006 | 0.4197 | 1.2441 | 0.9810 | 0.0409 | 0.2756 | 0.1192 |
| **Region 3** | 1152 | 0.2344 | 0.0609 | 0.9906 | 0.0499 | 0.2676 | 0.1093 |
| **Region 4** | 1193 | 0.2879 | 0.0516 | 0.9933 | 0.0468 | 0.2221 | 0.1325 |
| **Region 5** | 1142 | 0.4275 | 1.4396 | 0.9910 | 0.0401 | 0.2682 | 0.1899 |
| **Region 6** | 1012 | 0.2304 | 0.1554 | 0.9923 | 0.0546 | 0.2754 | 0.1048 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Region 7** | 972 | 0.3292 | 0.0001 | 0.9735 | 0.0832 | 0.2856 | 0.1114 |
| **Region 8** | 1235 | 0.3789 | 0.0516 | 0.9948 | 0.0420 | 0.2460 | 0.1374 |

Source: Author‘s calculations.

## Conclusion

Each inequality measure used above quantifies the inequality of the income distribution in comparison with a different theoretical probability model. Based on the results shown in Tab. 5 above, we came to the following conclusions. A correlation coefficients of 0.97 and higher signify that the Generalized Kappa distribution is sufficiently flexible for regional studies in Slovakia. V(inequal.) is the inequality measure calculated according to Tarsitano’s approach. The higher the V(inequal.) measure, the higher is the measured inequality. The Gini coefficient is another measure of inequality, which should correspond with the V(inequal.) measures. However, if we compare V(inequal.) and Gini of Regions 1 and 7, there is a discrepancy, which can be explained by the quality of the respective models. Finally, the results obtained using the De Vergottini’s method do not correspond with results obtained using V(inequal.) and Gini. Further research needs to be conducted to be able to further analyze the results obtained using the three methods mentioned above.

## References

1. BARTOŠOVÁ, J.- BÍNA, V.: *Mixture Models of Household Income Distribution in the Czech Republic*. Bratislava 06. 2.2007– 09.02.2007. In: 6th International Conference APLIMAT 2007, Part I. Bratislava : Slovak University of Technology, 2007, s. 307-316, ISBN 978-80-969562-4-1, 2007.
2. DAGUM, C.: *A Measure of Inequality Between Income Distributions*, Economie Apliquée XXXI, (3-4), 401-413, 1978.
3. DAGUM, C.: *Analyses of Icome Distributio and Inequality by Education and Sex in Canada*, Advances in Econometrics, volume 4, 167-227, 1985.
4. DAGUM, C.: *Income Distribution Models*, The Encyclopedia of Statistical Science, 4, 27-34, 1984.
5. DAGUM, C: *A new model of personal distribution: specification and estimation*, Economie Appliquée, Tome XXX, N. 3., 1977.
6. EDGEWORTH, F.Y.: *On the Representation of Statistics by Mathematical Formulae,* Journal of the Royal Statistical Society 61, Part I - December, 670-700, Part II - 62, March, 125-140, Part III – 62, June, 373-385, Part IV – 62, September, 534-555, 1898-1899.
7. FOWLKES, E. B.: *A Folio of Distributions, A collection of Theoretical Q-Q plots*, Marcel Deckker, New York, 1987.
8. GILCHRIST, W.G.: *Modelling with quantile distribution functions*, In: Journal of Applied Statistics, Vol.24, No. 1, 113-122, 1997.
9. GILCHRIST, W.G.*: Statistical modelling with quantile functions*, Chapman & Hall, 2000.

10. HOSKING J. R. M.: *The four-parameter kappa distribution*. IBM Journal of Research, Development, 38, 251-258, 1994.
11. CHAMPERNOWNE, D.G.,: *A Model of Income Distribution*, Economic Journal 63, June, 318-351, 1953.
12. CHIPMAN, J. S.: *The Theory and Measurement of Income Distribution*, Advances in Econometrics, 4, 135-165, 1985
13. RAMBERG, J. - SCHMEISER, B., , *An approximate method for generating asymmetric random variables*, Communications of the ACM, 17(2), ISSN 0001-0782, 1974.
14. SEN, A. K.: *On economic inequality*. Clarendon Press, Oxford, 1973.
15. PACÁKOVÁ, V. - SODOMOVÁ, E.: *Modelovanie pomocou kvantilov distribučných funkcií*. Ekonomika a informatika, No. 1, 30-44, 2003.
16. STANKOVIČOVÁ, I.: *Regionálne aspekty monetárnej chudoby na Slovensku,* Sociálny kapitál, ľudských kapitál a chudoba v regiónoch Slovenska, Košice : Ekonomická fakulta TU, 67-75, 2010.
17. TARSITANO, A.: *A new Q-Q plot and its application to income data*. Available on the internet: http://www3.unisi.it/eventi/GiniLorenz05/ABSTRACT_PAPER_24%20May/PAPER_Tarsitano.pdf
18. VOJTKOVÁ, M - LABUDOVÁ, V.: *Regionálna analýza príjmov a výdavkov v Slovenskej republike*, In: Ekonomický časopis, Vol.58, No. 8, 802-820, 2010.
19. ŽELINSKÝ, T: *Nerovnosť rozdeľovania príjmov v krajoch Slovenskej republiky*. In: Slovenská štatistika a demografia. Vol. 20, No. 1, 49-60, 2010.


## Acknowledgements

Research was partially supported and funded by the **VEGA Project No.01/0127/11**: The spatial distribution of poverty in the European Union.

## Contact


Ing. Ľubica Sipková, PhD.

University of Economics in Bratislava

Department of Statistics, Faculty of Economic Informatics

Dolnozemská 1, 852 35 Bratislava

Slovak Republic

e-mail: lubica.sipkova@euba.sk

Ass.Prof. Juraj Sipko, M.B.A., PhD.

Paneuropean University

Faculty of Economics and Business

Tematínska 10, 851 05 Bratislava 5

Slovak Republic

e-mail: juraj.sipko@uninova.sk